Mixing time and simulated annealing for the stochastic cellular automata, and beyond

Akira Sakai

Hokkaido University

January 13, 2022

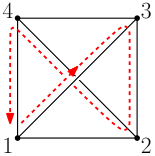
- Finding an optimal solution to a given problem $\overset{\text{equiv}}{\Longleftrightarrow}$ finding a ground state $\boldsymbol{\sigma}_{\text{GS}} \in \{\pm 1\}^V$ of the corresponding Ising Hamiltonian on a finite graph $G = (V, E)$:

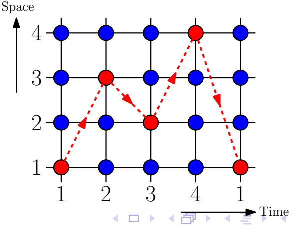
$$H(\boldsymbol{\sigma}) = -\frac{1}{2}\sum_{x,y\in V} J_{x,y}\sigma_x\sigma_y - \sum_{x\in V} h_x\sigma_x, \qquad H(\boldsymbol{\sigma}_{\text{GS}}) = \min_{\boldsymbol{\sigma}\in\{\pm1\}^V} H(\boldsymbol{\sigma}).$$

- E.g., the Traveling Salesman Problem (TSP):
  - $d_{x,y}$: a transpotation cost between x and y $(d_{x,x} = 0,\ d_{x,y} = d_{y,x})$.
  - $\xi_{(t,x)}$: 1 (or 0) if the salesman is (or isn't) at $x \in V$ at time $t \in \{1, \dots, |V|\}$.
  - Find a minimizer $\boldsymbol{\xi} = \{\xi_{(t,x)}\}$ of the cost function (where $c = \frac{1}{4}\|d\|_\infty$):

$$H(\boldsymbol{\xi}) = \sum_{x,y\in V} d_{x,y}\sum_{t=1}^{|V|}\xi_{(t,x)}\xi_{(t+1,y)} + c\left(\sum_{t=1}^{|V|}\left(\sum_{x\in V}\xi_{(t,x)} - 1\right)^2 + \sum_{x\in V}\left(\sum_{t=1}^{|V|}\xi_{(t,x)} - 1\right)^2\right).$$

  - Equivalent to finding a spin configuration $\boldsymbol{\sigma} = 2\boldsymbol{\xi} - \mathbf{1} \in \{-1, 1\}^V$ that minimizes the corresponding Hamiltonian.





- The Gibbs distribution:

$$\pi_\beta^{\mathrm{G}}(\boldsymbol{\sigma}) = \frac{e^{-\beta H(\boldsymbol{\sigma})}}{\sum_{\boldsymbol{\sigma}} e^{-\beta H(\boldsymbol{\sigma})}} \xrightarrow[\beta\uparrow\infty]{} \text{Unif(GSs)} = \frac{\mathbb{1}_{\{\boldsymbol{\sigma} \text{ is a GS}\}}}{\#\text{ of GSs}}.$$

- A standard Gibbs sampler (the Glauber dynamics):

$$(\boldsymbol{\sigma}^x)_y = \begin{cases} -\sigma_y & [y = x], \\ \sigma_y & [y \neq x], \end{cases} \qquad P_\beta^{\mathrm{G}}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \begin{cases} \frac{1}{|V|}\frac{e^{-\beta H(\boldsymbol{\sigma}^x)}}{e^{-\beta H(\boldsymbol{\sigma})} + e^{-\beta H(\boldsymbol{\sigma}^x)}} & [\boldsymbol{\eta} = \boldsymbol{\sigma}^x], \\ 1 - \sum_{x\in V} P^{\mathrm{G}}_{\boldsymbol{\sigma},\boldsymbol{\sigma}^x} & [\boldsymbol{\eta} = \boldsymbol{\sigma}], \\ 0 & [\text{o/w}]. \end{cases}$$

$$\Longrightarrow \ {}^\forall\boldsymbol{\sigma},\boldsymbol{\eta} \in \{\pm1\}^V : \underbrace{\pi_\beta^{\mathrm{G}}(\boldsymbol{\sigma})P_\beta^{\mathrm{G}}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \pi_\beta^{\mathrm{G}}(\boldsymbol{\eta})P_\beta^{\mathrm{G}}(\boldsymbol{\eta},\boldsymbol{\sigma})}_{\text{detailed balance}}, \quad {}^\forall\mu * (P_\beta^{\mathrm{G}})^{*n} \xrightarrow[n\uparrow\infty]{} \pi_\beta^{\mathrm{G}}.$$

- Simulated annealing : $\beta_n = {}^\exists c \log n \implies \mu * P^{\mathrm{G}}_{\beta_1} * \cdots * P^{\mathrm{G}}_{\beta_n} \xrightarrow[n\uparrow\infty]{} \text{Unif(GSs)}$.
  by Hajek (1988), Catoni (1992)
- In reality:
  - Stopping at $n \ll \infty$ without knowing how close $\mu * P^{\mathrm{G}}_{\beta_1} * \cdots * P^{\mathrm{G}}_{\beta_n}$ is to Unif(GSs).
  - Very slow, due to single-spin updates, log cooling schedule, convergence in total variation, etc.

- The Gibbs distribution:

$$\pi_\beta^{\mathrm{G}}(\boldsymbol{\sigma}) = \frac{e^{-\beta H(\boldsymbol{\sigma})}}{\sum_{\boldsymbol{\sigma}} e^{-\beta H(\boldsymbol{\sigma})}} \xrightarrow[\beta\uparrow\infty]{} \text{Unif(GSs)} = \frac{\mathbb{1}_{\{\boldsymbol{\sigma} \text{ is a GS}\}}}{\#\text{ of GSs}}.$$

- A standard Gibbs sampler (the Glauber dynamics):

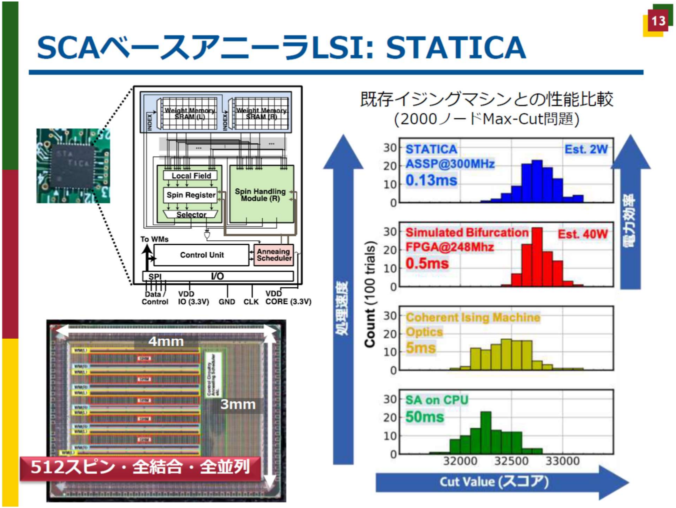
$$(\boldsymbol{\sigma}^x)_y = \begin{cases} -\sigma_y & [y = x], \\ \sigma_y & [y \neq x], \end{cases} \qquad P_\beta^{\mathrm{G}}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \begin{cases} \frac{1}{|V|} \frac{e^{-\beta H(\boldsymbol{\sigma}^x)}}{e^{-\beta H(\boldsymbol{\sigma})} + e^{-\beta H(\boldsymbol{\sigma}^x)}} & [\boldsymbol{\eta} = \boldsymbol{\sigma}^x], \\ 1 - \sum_{x\in V} P^{\mathrm{G}}_{\boldsymbol{\sigma},\boldsymbol{\sigma}^x} & [\boldsymbol{\eta} = \boldsymbol{\sigma}], \\ 0 & [\text{o/w}]. \end{cases}$$

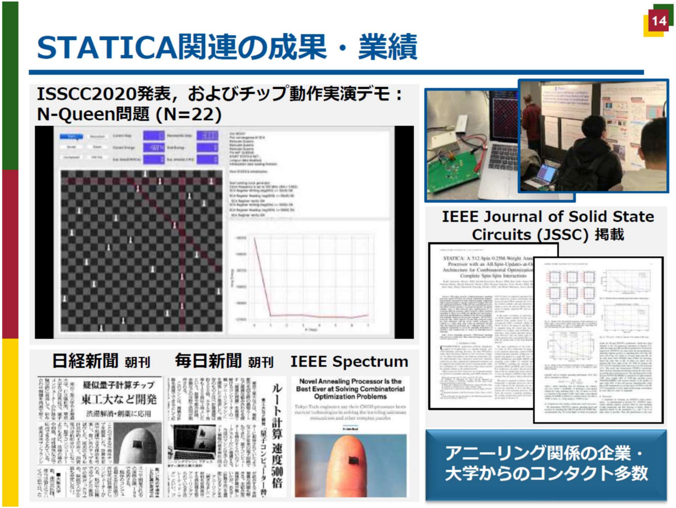
$$\Longrightarrow \ {}^\forall\boldsymbol{\sigma},\boldsymbol{\eta} \in \{\pm 1\}^V : \underbrace{\pi_\beta^{\mathrm{G}}(\boldsymbol{\sigma})P_\beta^{\mathrm{G}}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \pi_\beta^{\mathrm{G}}(\boldsymbol{\eta})P_\beta^{\mathrm{G}}(\boldsymbol{\eta},\boldsymbol{\sigma})}_{\text{detailed balance}}, \quad {}^\forall\mu * (P_\beta^{\mathrm{G}})^{*n} \xrightarrow[n\uparrow\infty]{} \pi_\beta^{\mathrm{G}}.$$

- Simulated annealing : $\beta_n = {}^\exists c \log n \implies \mu * P^{\mathrm{G}}_{\beta_1} * \cdots * P^{\mathrm{G}}_{\beta_n} \xrightarrow[n\uparrow\infty]{} \text{Unif(GSs)}.$
  by Hajek (1988), Catoni (1992)
- In reality:
  - Stopping at $n \ll \infty$ without knowing how close $\mu * P^{\mathrm{G}}_{\beta_1} * \cdots * P^{\mathrm{G}}_{\beta_n}$ is to Unif(GSs).
  - Very slow, due to single-spin updates, log cooling schedule, convergence in total variation, etc.

- What we want:
  Faster dynamics based on multi-spin updates.
- What we propose:
  Use the stochastic cellular automata (SCA) inspired by Dai Pra et al. (2012).
- What we have proven:
  - Mixing is much faster than Glauber in the high-temperature regime.
  - A standard cooling schedule works for the SCA, too, to find a $\boldsymbol{\sigma}_{\text{GS}}$.
- Ongoing work (supported by numerical evidence):
  - ε-SCA, where the pinning parameters $\boldsymbol{q} = \{q_x\}_{x \in V}$ are turned off and a collection of spins is chosen by Binom$(|V|, \varepsilon)$.
  - First hit to the target states, under an exponential cooling schedule.
    (First hit to the target states $\leftrightarrow$ convergence in total variation; exponential $\leftrightarrow$ logarithmic)

- What we want:
  Faster dynamics based on multi-spin updates.
- What we propose:
  Use the stochastic cellular automata (SCA) inspired by Dai Pra et al. (2012).
- What we have proven:
  - Mixing is much faster than Glauber in the high-temperature regime.
  - A standard cooling schedule works for the SCA, too, to find a $\boldsymbol{\sigma}_{\mathrm{GS}}$.
- Ongoing work (supported by numerical evidence):
  - ε-SCA, where the pinning parameters $\boldsymbol{q} = \{q_x\}_{x \in V}$ are turned off and a collection of spins is chosen by $\mathrm{Binom}(|V|, \varepsilon)$.
  - $\underbrace{\text{First hit to the target states}}_{\leftrightarrow \text{ convergence in total variation}}$, under an $\underbrace{\text{exponential}}_{\leftrightarrow \text{ logarithmic}}$ cooling schedule.


13
SCAベースアニーラLSI: STATICA
Weight Memory SRAM (L)
Weight Memory SRAM (R)
INDEX
Local Field
Spin Register
Selector
Spin Handling Module (R)
To WMs
Control Unit
Anneaing Scheduler
SPI
I/O
Data / Control
VDD IO (3.3V)
GND
CLK
VDD CORE (3.3V)
4mm
3mm
512スピン・全結合・全並列
既存イジングマシンとの性能比較
(2000ノードMax-Cut問題)
処理速度
電力効率
Count (100 trials)
STATICA
ASSP@300MHz
0.13ms
Est. 2W
Simulated Bifurcation
FPGA@248Mhz
0.5ms
Est. 40W
Coherent Ising Machine
Optics
5ms
SA on CPU
50ms
32000
32500
33000
Cut Value (スコア)


14
STATICA関連の成果・業績
ISSCC2020発表，およびチップ動作実演デモ：
N-Queen問題 (N=22)
IEEE Journal of Solid State Circuits (JSSC) 掲載
日経新聞 朝刊
毎日新聞 朝刊
IEEE Spectrum
疑似量子計算チップ
東工大など開発
渋滞解消・創薬に応用
Novel Annealing Processor Is the Best Ever at Solving Combinatorial Optimization Problems
アニーリング関係の企業・大学からのコンタクト多数

- The doubled Hamiltonian and the cavity field (generalization of Dai Pra et al. (2012)):

$$\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\eta}) = -\frac{1}{2}\sum_{x,y} J_{x,y}\sigma_x\eta_y - \frac{1}{2}\sum_x h_x(\sigma_x+\eta_x) - \frac{1}{2}\sum_x q_x\sigma_x\eta_x$$
$$= -\frac{1}{2}\sum_x h_x\sigma_x - \frac{1}{2}\sum_x \Big(\underbrace{\sum_y J_{x,y}\sigma_y + h_x}_{\tilde{h}_x(\boldsymbol{\sigma})} + q_x\sigma_x\Big)\eta_x .$$

- The SCA and its equilibrium distribution:

$$w_{\beta,q}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\eta}} e^{-\beta\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\eta})} = \prod_x e^{\frac{\beta}{2}h_x\sigma_x} 2\cosh\Big(\tfrac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x)\Big),$$

$$P^{\mathrm{SCA}}_{\beta,q}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \frac{e^{-\beta\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\eta})}}{w_{\beta,q}(\boldsymbol{\sigma})} = \underbrace{\prod_x \frac{e^{\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x)\eta_x}}{2\cosh(\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x))}}_{\text{multi-spin simultaneous update}} = \prod_x \frac{1+\eta_x\tanh(\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x))}{2}$$

$$\pi^{\mathrm{SCA}}_{\beta,q}(\boldsymbol{\sigma}) = \frac{w_{\beta,q}(\boldsymbol{\sigma})}{\sum_{\boldsymbol{\sigma}} w_{\beta,q}(\boldsymbol{\sigma})} \xrightarrow[\min q_x\uparrow\infty]{} \pi^{\mathrm{G}}_{\beta}(\boldsymbol{\sigma}) \qquad \Big[\because\ \tilde{H}(\boldsymbol{\sigma},\boldsymbol{\sigma}) = H(\boldsymbol{\sigma}) - \tfrac{1}{2}\textstyle\sum_x q_x\Big]$$

$$\Longrightarrow\ {}^\forall\boldsymbol{\sigma},\boldsymbol{\eta}\in\{\pm1\}^V :\ \underbrace{\pi^{\mathrm{SCA}}_{\beta,q}(\boldsymbol{\sigma})P^{\mathrm{SCA}}_{\beta,q}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \pi^{\mathrm{SCA}}_{\beta,q}(\boldsymbol{\eta})P^{\mathrm{SCA}}_{\beta,q}(\boldsymbol{\eta},\boldsymbol{\sigma})}_{\text{detailed balance}},\quad {}^\forall\mu * (P^{\mathrm{SCA}}_{\beta,q})^{*n} \xrightarrow[n\uparrow\infty]{} \pi^{\mathrm{SCA}}_{\beta,q}.$$

- The doubled Hamiltonian and the cavity field (generalization of Dai Pra et al. (2012)):

$$\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\eta}) = -\frac{1}{2}\sum_{x,y} J_{x,y}\sigma_x\eta_y - \frac{1}{2}\sum_x h_x(\sigma_x+\eta_x) - \frac{1}{2}\sum_x q_x\sigma_x\eta_x$$
$$= -\frac{1}{2}\sum_x h_x\sigma_x - \frac{1}{2}\sum_x \Big(\underbrace{\sum_y J_{x,y}\sigma_y + h_x}_{\tilde{h}_x(\boldsymbol{\sigma})} + q_x\sigma_x\Big)\eta_x.$$

- The SCA and its equilibrium distribution:

$$w_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\eta}} e^{-\beta\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\eta})} = \prod_x e^{\frac{\beta}{2}h_x\sigma_x} 2\cosh\Big(\tfrac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x)\Big),$$

$$P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \frac{e^{-\beta\tilde{H}(\boldsymbol{\sigma},\boldsymbol{\eta})}}{w_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma})} = \underbrace{\prod_x \frac{e^{\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x)\eta_x}}{2\cosh(\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x))}}_{\text{multi-spin simultaneous update}} = \prod_x \frac{1+\eta_x\tanh(\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x))}{2}$$

$$\pi^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma}) = \frac{w_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma})}{\sum_{\boldsymbol{\sigma}} w_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma})} \xrightarrow[\min q_x\uparrow\infty]{} \pi^{\mathrm{G}}_\beta(\boldsymbol{\sigma}) \qquad \Big[\because \tilde{H}(\boldsymbol{\sigma},\boldsymbol{\sigma}) = H(\boldsymbol{\sigma}) - \tfrac{1}{2}\textstyle\sum_x q_x\Big]$$

$$\Longrightarrow \ {}^\forall\boldsymbol{\sigma},\boldsymbol{\eta}\in\{\pm1\}^V : \ \underbrace{\pi^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma})P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \pi^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\eta})P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\eta},\boldsymbol{\sigma})}_{\text{detailed balance}}, \quad {}^\forall\mu * (P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}})^{*n} \xrightarrow[n\uparrow\infty]{} \pi^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}.$$

Proposition 1 (Okuyama, Sonobe, Kawarabayashi and Yamaoka, *PRE* 100 (2019))

If $\min_x q_x \geq \frac{\lambda}{2}$*, where* λ *is the largest eigenvalue of the matrix* $[-J_{x,y}]_{V\times V}$*, then*

$$\min_{\sigma,\eta} \tilde{H}(\sigma,\eta) = \min_{\sigma} \tilde{H}(\sigma,\sigma), \qquad \arg\min_{\sigma} \tilde{H}(\sigma,\sigma) = GS.$$

N.b., q_x does not have to be so large, if the minimum is taken over a smaller set of spin configurations (Kawamoto).

Theorem 2 (with Fukushima-Kimura, Handa, Kamakura, Kamijima, Kawamura (2021))

If

$$r = \max_x \left(\tanh \frac{\beta q_x}{2} + \sum_y \tanh \frac{\beta |J_{x,y}|}{2} \right) < 1,$$

then, for any $\varepsilon > 0$,

$$T^{\mathrm{SCA}}_{\mathrm{mix}}(\varepsilon) = \min\left\{ n : \max_{\sigma} \left\| \delta_\sigma * (P^{\mathrm{SCA}}_{\beta,q})^n - \pi^{\mathrm{SCA}}_{\beta,q} \right\|_{\mathrm{TV}} \leq \varepsilon \right\} \leq \frac{\log|V| + \log(1/\varepsilon)}{\log(1/r)}.$$

C.f., Levin, Peres, Wilmer (2008) for the Glauber dynamics:

$$T^{\mathrm{G}}_{\mathrm{mix}}(\varepsilon) \geq \left(\frac{|V|}{2} - 1 \right) \log \frac{1}{2\varepsilon}.$$

Proposition 1 (Okuyama, Sonobe, Kawarabayashi and Yamaoka, *PRE* 100 (2019))

If $\min_x q_x \geq \frac{\lambda}{2}$, *where* λ *is the largest eigenvalue of the matrix* $[-J_{x,y}]_{V\times V}$, *then*

$$\min_{\sigma,\eta} \tilde{H}(\sigma,\eta) = \min_{\sigma} \tilde{H}(\sigma,\sigma), \qquad \arg\min_{\sigma} \tilde{H}(\sigma,\sigma) = GS.$$

N.b., q_x does not have to be so large, if the minimum is taken over a smaller set of spin configurations (Kawamoto).

Theorem 2 (with Fukushima-Kimura, Handa, Kamakura, Kamijima, Kawamura (2021))

If

$$r = \max_x \left(\tanh \frac{\beta q_x}{2} + \sum_y \tanh \frac{\beta |J_{x,y}|}{2} \right) < 1,$$

then, for any $\varepsilon > 0$,

$$T^{\mathrm{SCA}}_{\mathrm{mix}}(\varepsilon) = \min\left\{ n : \max_{\sigma} \left\| \delta_\sigma * (P^{\mathrm{SCA}}_{\beta,q})^n - \pi^{\mathrm{SCA}}_{\beta,q} \right\|_{\mathrm{TV}} \leq \varepsilon \right\} \leq \frac{\log|V| + \log(1/\varepsilon)}{\log(1/r)}.$$

C.f., Levin, Peres, Wilmer (2008) for the Glauber dynamics:

$$T^{\mathrm{G}}_{\mathrm{mix}}(\varepsilon) \geq \left(\frac{|V|}{2} - 1 \right) \log \frac{1}{2\varepsilon}.$$

Idea of the proof of Theorem 2:

- Suppose that the transportation metric, defined by

$$\rho_{\mathrm{TM}}(\mu,\nu)=\inf_{(X,Y):\text{coupling}}\mathbb{E}_{\mu,\nu}\Big[\sum_y \mathbb{1}_{\{X_y\neq Y_y\}}\Big] \qquad \big(\text{n.b., } \mathbb{E}_{\mu,\nu}[\mathbb{1}_{\{X\neq Y\}}]\geq\|\mu-\nu\|_{\mathrm{TV}}\big),$$

satisfies

$$\rho_{\mathrm{TM}}\Big(\delta_{\boldsymbol{\sigma}}*P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}},\ \delta_{\boldsymbol{\sigma}^x}*P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}\Big)\leq r. \tag{1}$$

- Then, by repeated use of the triangle inequality,

$$\rho_{\mathrm{TM}}\Big(\delta_{\boldsymbol{\sigma}}*P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}},\ \delta_{\boldsymbol{\eta}}*P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}\Big)\leq\cdots\leq r\sum_y \mathbb{1}_{\{\sigma_y\neq\eta_y\}}\leq|V|r.$$

Consequently

$$\begin{aligned}\Big\|\delta_{\boldsymbol{\sigma}}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}})^{*n}-\pi^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}\Big\|_{\mathrm{TV}}&\leq\rho_{\mathrm{TM}}\Big(\delta_{\boldsymbol{\sigma}}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}})^{*n},\ \pi^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}\Big)\\&=\rho_{\mathrm{TM}}\Big(\delta_{\boldsymbol{\sigma}}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}})^{*n},\ \pi^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}})^{*n}\Big)\\&\leq\sum_{\boldsymbol{\eta}}\pi^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\eta})\ \rho_{\mathrm{TM}}\Big(\delta_{\boldsymbol{\sigma}}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}})^{*n},\ \delta_{\boldsymbol{\eta}}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}})^{*n}\Big)\leq|V|r^n.\end{aligned}$$

Idea of the proof of Theorem 2:

- Suppose that the transportation metric, defined by

$$\rho_{\mathrm{TM}}(\mu,\nu)=\inf_{(X,Y):\text{coupling}}\mathbb{E}_{\mu,\nu}\Big[\sum_y \mathbb{1}_{\{X_y\neq Y_y\}}\Big] \qquad \Big(\text{n.b., } \mathbb{E}_{\mu,\nu}[\mathbb{1}_{\{X\neq Y\}}]\geq\|\mu-\nu\|_{\mathrm{TV}}\Big),$$

satisfies

$$\rho_{\mathrm{TM}}\Big(\delta_{\boldsymbol\sigma}*P^{\mathrm{SCA}}_{\beta,\boldsymbol q},\ \delta_{\boldsymbol\sigma^x}*P^{\mathrm{SCA}}_{\beta,\boldsymbol q}\Big)\leq r. \tag{1}$$

- Then, by repeated use of the triangle inequality,

$$\rho_{\mathrm{TM}}\Big(\delta_{\boldsymbol\sigma}*P^{\mathrm{SCA}}_{\beta,\boldsymbol q},\ \delta_{\boldsymbol\eta}*P^{\mathrm{SCA}}_{\beta,\boldsymbol q}\Big)\leq\cdots\leq r\sum_y \mathbb{1}_{\{\sigma_y\neq\eta_y\}}\leq|V|r.$$

Consequently

$$\begin{aligned}\Big\|\delta_{\boldsymbol\sigma}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol q})^{*n}-\pi^{\mathrm{SCA}}_{\beta,\boldsymbol q}\Big\|_{\mathrm{TV}}&\leq\rho_{\mathrm{TM}}\Big(\delta_{\boldsymbol\sigma}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol q})^{*n},\ \pi^{\mathrm{SCA}}_{\beta,\boldsymbol q}\Big)\\&=\rho_{\mathrm{TM}}\Big(\delta_{\boldsymbol\sigma}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol q})^{*n},\ \pi^{\mathrm{SCA}}_{\beta,\boldsymbol q}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol q})^{*n}\Big)\\&\leq\sum_{\boldsymbol\eta}\pi^{\mathrm{SCA}}_{\beta,\boldsymbol q}(\boldsymbol\eta)\ \rho_{\mathrm{TM}}\Big(\delta_{\boldsymbol\sigma}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol q})^{*n},\ \delta_{\boldsymbol\eta}*(P^{\mathrm{SCA}}_{\beta,\boldsymbol q})^{*n}\Big)\leq|V|r^n.\end{aligned}$$

- To show (1), use the $[0,1]$-uniform random variables $\{U_y\}_{y\in V}$ and define

$$\begin{cases} X_y = 2\mathbb{1}_{\{U_y\le p_y^+(\boldsymbol{\sigma})\}} - 1, \\ Y_y = 2\mathbb{1}_{\{U_y\le p_y^+(\boldsymbol{\sigma}^x)\}} - 1, \end{cases} \qquad \text{where} \quad p_y^+(\boldsymbol{\sigma}) = \frac{1+\tanh(\frac{\beta}{2}(\tilde{h}_y(\boldsymbol{\sigma})+q_y\sigma_y))}{2}.$$

- Since $|\tanh(a+b)-\tanh(a-b)| \le 2\tanh|b|$ holds for any $a,b\in\mathbb{R}$,

$$|p_x^+(\boldsymbol{\sigma})-p_x^+(\boldsymbol{\sigma}^x)| \le \tanh\frac{\beta q_x}{2}, \qquad \sum_{y\sim x}|p_y^+(\boldsymbol{\sigma})-p_y^+(\boldsymbol{\sigma}^x)| \le \sum_y \tanh\frac{\beta|J_{x,y}|}{2}.$$

- Therefore

$$\begin{aligned}\mathbb{E}_{\delta_{\sigma},\delta_{\sigma^x}}\Big[\sum_y \mathbb{1}_{\{X_y\ne Y_y\}}\Big] &= \sum_y \mathbb{P}_{\delta_{\sigma},\delta_{\sigma^x}}(X_y\ne Y_y) \\ &= \sum_y |p_y^+(\boldsymbol{\sigma})-p_y^+(\boldsymbol{\sigma}^x)| \\ &= |p_x^+(\boldsymbol{\sigma})-p_x^+(\boldsymbol{\sigma}^x)| + \sum_{y\sim x}|p_y^+(\boldsymbol{\sigma})-p_y^+(\boldsymbol{\sigma}^x)| \\ &\le \max_x\Big(\tanh\frac{\beta q_x}{2} + \sum_y \tanh\frac{\beta|J_{x,y}|}{2}\Big) = r.\end{aligned}$$

- To show (1), use the $[0,1]$-uniform random variables $\{U_y\}_{y\in V}$ and define

$$\begin{cases} X_y = 2\mathbb{1}_{\{U_y \le p_y^+(\boldsymbol{\sigma})\}} - 1, \\ Y_y = 2\mathbb{1}_{\{U_y \le p_y^+(\boldsymbol{\sigma}^x)\}} - 1, \end{cases} \qquad \text{where} \quad p_y^+(\boldsymbol{\sigma}) = \frac{1 + \tanh(\frac{\beta}{2}(\tilde{h}_y(\boldsymbol{\sigma}) + q_y\sigma_y))}{2}.$$

- Since $|\tanh(a+b) - \tanh(a-b)| \le 2\tanh|b|$ holds for any $a, b \in \mathbb{R}$,

$$|p_x^+(\boldsymbol{\sigma}) - p_x^+(\boldsymbol{\sigma}^x)| \le \tanh\frac{\beta q_x}{2}, \qquad \sum_{y\sim x} |p_y^+(\boldsymbol{\sigma}) - p_y^+(\boldsymbol{\sigma}^x)| \le \sum_y \tanh\frac{\beta|J_{x,y}|}{2}.$$

- Therefore

$$\begin{aligned} \mathbb{E}_{\delta_{\boldsymbol{\sigma}},\delta_{\boldsymbol{\sigma}^x}}\Big[\sum_y \mathbb{1}_{\{X_y \ne Y_y\}}\Big] &= \sum_y \mathbb{P}_{\delta_{\boldsymbol{\sigma}},\delta_{\boldsymbol{\sigma}^x}}(X_y \ne Y_y) \\ &= \sum_y |p_y^+(\boldsymbol{\sigma}) - p_y^+(\boldsymbol{\sigma}^x)| \\ &= |p_x^+(\boldsymbol{\sigma}) - p_x^+(\boldsymbol{\sigma}^x)| + \sum_{y\sim x} |p_y^+(\boldsymbol{\sigma}) - p_y^+(\boldsymbol{\sigma}^x)| \\ &\le \max_x \bigg(\tanh\frac{\beta q_x}{2} + \sum_y \tanh\frac{\beta|J_{x,y}|}{2}\bigg) = r. \end{aligned}$$

Simulated annealing for the SCA:

Theorem 3 (with Fukushima-Kimura, Handa, Kamakura, Kamijima, Kawamura (2021))

$$\Gamma_x = q_x + |h_x| + \sum_y |J_{x,y}|, \quad \beta_n = \frac{\log n}{\sum_x \Gamma_x}, \quad P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} \overset{j\le n}{\equiv} P^{\rm SCA}_{\beta_j,\boldsymbol{q}} * P^{\rm SCA}_{\beta_{j+1},\boldsymbol{q}} * \cdots * P^{\rm SCA}_{\beta_n,\boldsymbol{q}}$$

$$\Rightarrow \quad {}^\forall j \in \mathbb{N}, \quad \sup_\mu \left\| \mu * P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} - \textit{Unif(GSs)} \right\|_{\rm TV} \xrightarrow[n\uparrow\infty]{} 0. \tag{2}$$

N.b., $\beta_n \propto \log n$ for a single-spin flip MCMC (the Metropolis sampler).

Idea of the proof of Theorem 3:

- According to Brémaud (1999), a sufficient condition for the strong ergodicity (2) is
  - (i) the weak ergodicity of $\{P^{\rm SCA}_{\beta_n,\boldsymbol{q}}\}_{n\in\mathbb{N}}$: $\sup_{\mu,\nu} \left\| \mu * P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} - \nu * P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} \right\|_{\rm TV} \xrightarrow[n\uparrow\infty]{} 0.$
  - (ii) $\sum_{n=1}^\infty \left\| \pi^{\rm SCA}_{\beta_{n+1},\boldsymbol{q}} - \pi^{\rm SCA}_{\beta_n,\boldsymbol{q}} \right\|_{\rm TV} < \infty.$

$$\therefore \ \|\mu * P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} - \pi^{\rm SCA}_{\infty,\boldsymbol{q}}\|_{\rm TV} = \|\mu * (P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} - \Pi^{\rm SCA}_{\infty,\boldsymbol{q}})\|_{\rm TV}$$

$$\le \underbrace{\|\mu * (P^{\rm SCA}_{[\beta_j,\beta_\ell],\boldsymbol{q}} - \Pi^{\rm SCA}_{\beta_\ell,\boldsymbol{q}}) P^{\rm SCA}_{[\beta_\ell,\beta_n],\boldsymbol{q}}\|_{\rm TV}}_{\le 2\delta\left(P^{\rm SCA}_{[\beta_\ell,\beta_n],\boldsymbol{q}}\right):\ \text{Dobrushin's ergodic coeff.} \ \longrightarrow 0 \ \because \text{(i)}} + \underbrace{\|\mu * (\Pi^{\rm SCA}_{\beta_\ell,\boldsymbol{q}} P^{\rm SCA}_{[\beta_\ell,\beta_n],\boldsymbol{q}} - \Pi^{\rm SCA}_{\beta_n,\boldsymbol{q}})\|_{\rm TV}}_{\le \sum_{i=\ell+1}^n \|\mu*(\Pi^{\rm SCA}_{\beta_{i-1},\boldsymbol{q}} - \Pi^{\rm SCA}_{\beta_i,\boldsymbol{q}}) P^{\rm SCA}_{[\beta_i,\beta_n],\boldsymbol{q}}\|_{\rm TV} \le \sum_{i=\ell+1}^n \|\pi^{\rm SCA}_{\beta_{i-1},\boldsymbol{q}} - \pi^{\rm SCA}_{\beta_i,\boldsymbol{q}}\|_{\rm TV} \longrightarrow 0 \ \because \text{(ii)}} + \underbrace{\|\mu * (\Pi^{\rm SCA}_{\beta_n,\boldsymbol{q}} - \Pi^{\rm G}_{\infty})\|_{\rm TV}}_{\longrightarrow 0 \ \because \min q_x \ge \frac{\lambda}{2}}.$$

Simulated annealing for the SCA:

Theorem 3 (with Fukushima-Kimura, Handa, Kamakura, Kamijima, Kawamura (2021))

$$\Gamma_x = q_x + |h_x| + \sum_y |J_{x,y}|, \quad \beta_n = \frac{\log n}{\sum_x \Gamma_x}, \quad P^{\text{SCA}}_{[\beta_j,\beta_n],\boldsymbol{q}} \overset{j\le n}{=} P^{\text{SCA}}_{\beta_j,\boldsymbol{q}} * P^{\text{SCA}}_{\beta_{j+1},\boldsymbol{q}} * \cdots * P^{\text{SCA}}_{\beta_n,\boldsymbol{q}}$$

$$\Rightarrow \quad {}^\forall j \in \mathbb{N}, \quad \sup_\mu \left\| \mu * P^{\text{SCA}}_{[\beta_j,\beta_n],\boldsymbol{q}} - \mathit{Unif(GSs)} \right\|_{\text{TV}} \xrightarrow[n\uparrow\infty]{} 0. \tag{2}$$

N.b., $\beta_n \propto \log n$ for a single-spin flip MCMC (the Metropolis sampler).

Idea of the proof of Theorem 3:

- According to Brémaud (1999), a sufficient condition for the strong ergodicity (2) is
  - (i) the weak ergodicity of $\{P^{\text{SCA}}_{\beta_n,\boldsymbol{q}}\}_{n\in\mathbb{N}}$: $\sup_{\mu,\nu} \left\| \mu * P^{\text{SCA}}_{[\beta_j,\beta_n],\boldsymbol{q}} - \nu * P^{\text{SCA}}_{[\beta_j,\beta_n],\boldsymbol{q}} \right\|_{\text{TV}} \xrightarrow[n\uparrow\infty]{} 0.$
  - (ii) $\sum_{n=1}^\infty \left\| \pi^{\text{SCA}}_{\beta_{n+1},\boldsymbol{q}} - \pi^{\text{SCA}}_{\beta_n,\boldsymbol{q}} \right\|_{\text{TV}} < \infty.$

$$\therefore \left\| \mu * P^{\text{SCA}}_{[\beta_j,\beta_n],\boldsymbol{q}} - \pi^{\text{SCA}}_{\infty,\boldsymbol{q}} \right\|_{\text{TV}} = \left\| \mu * (P^{\text{SCA}}_{[\beta_j,\beta_n],\boldsymbol{q}} - \Pi^{\text{SCA}}_{\infty,\boldsymbol{q}}) \right\|_{\text{TV}}$$

$$\le \underbrace{\left\| \mu * (P^{\text{SCA}}_{[\beta_j,\beta_\ell],\boldsymbol{q}} - \Pi^{\text{SCA}}_{\beta_\ell,\boldsymbol{q}}) P^{\text{SCA}}_{[\beta_\ell,\beta_n],\boldsymbol{q}} \right\|_{\text{TV}}}_{\le 2\delta(P^{\text{SCA}}_{[\beta_\ell,\beta_n],\boldsymbol{q}}):\ \text{Dobrushin's ergodic coeff.} \ \longrightarrow 0 \ \because (i)} + \underbrace{\left\| \mu * (\Pi^{\text{SCA}}_{\beta_\ell,\boldsymbol{q}} P^{\text{SCA}}_{[\beta_\ell,\beta_n],\boldsymbol{q}} - \Pi^{\text{SCA}}_{\beta_n,\boldsymbol{q}}) \right\|_{\text{TV}}}_{\le \sum_{i=\ell+1}^n \left\| \mu * (\Pi^{\text{SCA}}_{\beta_{i-1},\boldsymbol{q}} - \Pi^{\text{SCA}}_{\beta_i,\boldsymbol{q}}) P^{\text{SCA}}_{[\beta_i,\beta_n],\boldsymbol{q}} \right\|_{\text{TV}}} + \underbrace{\left\| \mu * (\Pi^{\text{SCA}}_{\beta_n,\boldsymbol{q}} - \Pi^{\text{G}}_\infty) \right\|_{\text{TV}}}_{\longrightarrow 0 \ \because \min q_x \ge \frac{\lambda}{2}}.$$

$$\le \sum_{i=\ell+1}^n \left\| \pi^{\text{SCA}}_{\beta_{i-1},\boldsymbol{q}} - \pi^{\text{SCA}}_{\beta_i,\boldsymbol{q}} \right\|_{\text{TV}} \longrightarrow 0 \ \because (ii)$$

Simulated annealing for the SCA:

Theorem 3 (with Fukushima-Kimura, Handa, Kamakura, Kamijima, Kawamura (2021))

$$\Gamma_x = q_x + |h_x| + \sum_y |J_{x,y}|, \quad \beta_n = \frac{\log n}{\sum_x \Gamma_x}, \quad P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} \overset{j\le n}{=} P^{\rm SCA}_{\beta_j,\boldsymbol{q}} * P^{\rm SCA}_{\beta_{j+1},\boldsymbol{q}} * \cdots * P^{\rm SCA}_{\beta_n,\boldsymbol{q}}$$

$$\Rightarrow \quad {}^\forall j \in \mathbb{N}, \quad \sup_\mu \left\| \mu * P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} - \textit{Unif(GSs)} \right\|_{\rm TV} \xrightarrow[n\uparrow\infty]{} 0. \tag{2}$$

N.b., $\beta_n \propto \log n$ for a single-spin flip MCMC (the Metropolis sampler).

Idea of the proof of Theorem 3:

- According to Brémaud (1999), a sufficient condition for the strong ergodicity (2) is
  - (i) the weak ergodicity of $\{P^{\rm SCA}_{\beta_n,\boldsymbol{q}}\}_{n\in\mathbb{N}}$: $\sup_{\mu,\nu} \left\| \mu * P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} - \nu * P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} \right\|_{\rm TV} \xrightarrow[n\uparrow\infty]{} 0.$
  - (ii) $\sum_{n=1}^\infty \left\| \pi^{\rm SCA}_{\beta_{n+1},\boldsymbol{q}} - \pi^{\rm SCA}_{\beta_n,\boldsymbol{q}} \right\|_{\rm TV} < \infty.$

$$\therefore \ \|\mu * P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} - \pi^{\rm SCA}_{\infty,\boldsymbol{q}}\|_{\rm TV} = \|\mu * (P^{\rm SCA}_{[\beta_j,\beta_n],\boldsymbol{q}} - \Pi^{\rm SCA}_{\infty,\boldsymbol{q}})\|_{\rm TV}$$

$$\le \underbrace{\|\mu * (P^{\rm SCA}_{[\beta_j,\beta_\ell],\boldsymbol{q}} - \Pi^{\rm SCA}_{\beta_\ell,\boldsymbol{q}}) P^{\rm SCA}_{[\beta_\ell,\beta_n],\boldsymbol{q}}\|_{\rm TV}}_{\le 2\delta\left(P^{\rm SCA}_{[\beta_\ell,\beta_n],\boldsymbol{q}}\right)\text{: Dobrushin's ergodic coeff.} \ \longrightarrow 0 \ \because\text{(i)}} + \underbrace{\|\mu * (\Pi^{\rm SCA}_{\beta_\ell,\boldsymbol{q}} P^{\rm SCA}_{[\beta_\ell,\beta_n],\boldsymbol{q}} - \Pi^{\rm SCA}_{\beta_n,\boldsymbol{q}})\|_{\rm TV}} + \underbrace{\|\mu * (\Pi^{\rm SCA}_{\beta_n,\boldsymbol{q}} - \Pi^{\rm G}_\infty)\|_{\rm TV}}_{\longrightarrow 0 \ \because \min q_x \ge \frac{\lambda}{2}}.$$

$$\le \sum_{i=\ell+1}^n \|\mu * (\Pi^{\rm SCA}_{\beta_{i-1},\boldsymbol{q}} - \Pi^{\rm SCA}_{\beta_i,\boldsymbol{q}}) P^{\rm SCA}_{[\beta_i,\beta_n],\boldsymbol{q}}\|_{\rm TV}$$

$$\le \sum_{i=\ell+1}^n \|\pi^{\rm SCA}_{\beta_{i-1},\boldsymbol{q}} - \pi^{\rm SCA}_{\beta_i,\boldsymbol{q}}\|_{\rm TV} \longrightarrow 0 \ \because\text{(ii)}$$

- Dobrushin's ergodic coefficient:

$$\delta(P)=\max_{\sigma,\eta}\big\|\delta_\sigma*P-\delta_\eta*P\big\|_{\mathrm{TV}}=1-\min_{\sigma,\eta}\sum_{\tau}P(\sigma,\tau)\wedge P(\eta,\tau)$$

- Block criterion of weak ergodicity:

$$\sum_{n=1}^{\infty}\Big(1-\delta(P^{\mathrm{SCA}}_{\beta_n,q})\Big)=\infty \qquad\Rightarrow\qquad \delta(P^{\mathrm{SCA}}_{[\beta_\ell,\beta_n],q})\xrightarrow[n\uparrow\infty]{}0.$$

- The left is easy to show because

$$P^{\mathrm{SCA}}_{\beta_n,q}(\sigma,\tau)\ge\prod_x\frac{1}{1+e^{\beta_n|\tilde h_x(\sigma)+q_x\sigma_x|}}\ge\prod_x\frac{e^{-\beta_n\Gamma_x}}{2}=\frac{n^{-1}}{2^{|V|}}.$$

- For (ii), let $m=\min_{\sigma,\eta}\tilde H(\sigma,\eta)$ and

$$\mu_{\beta,q}(\sigma,\eta)=\frac{e^{-\beta\tilde H(\sigma,\eta)}}{\sum_{\sigma,\eta}e^{-\beta\tilde H(\sigma,\eta)}}\overset{\min q_x\ge\frac{\lambda}{2}}{=}\frac{e^{-\beta(\tilde H(\sigma,\eta)-m)}}{|\mathrm{GS}|+\sum_{\sigma,\eta:\tilde H(\sigma,\eta)>m}e^{-\beta(\tilde H(\sigma,\eta)-m)}}\xrightarrow[\beta\uparrow\infty]{}\pi^{\mathrm{G}}_\infty(\sigma)\,\delta_{\sigma,\eta}.$$

Since
$$\frac{\partial\mu_{\beta,q}(\sigma,\eta)}{\partial\beta}=\Big(\underbrace{\mathbb{E}_{\mu_{\beta,q}}[\tilde H]}_{\xrightarrow[\beta\uparrow\infty]{}m}-\tilde H(\sigma,\eta)\Big)\mu_{\beta,q}(\sigma,\eta)\begin{cases}>0 & \forall\beta<\infty\ \text{ if }\ \tilde H(\sigma,\eta)=m,\\ <0 & \text{if }\ \beta\gg1\ \&\ \tilde H(\sigma,\eta)>m,\end{cases}$$

we can show $\sum_{n=N}^{\infty}\|\pi^{\mathrm{SCA}}_{\beta_{n+1}}-\pi^{\mathrm{SCA}}_{\beta_n}\|_{\mathrm{TV}}\le\frac{3}{2}$ for $N\gg1$

- Dobrushin's ergodic coefficient:

$$\delta(P)=\max_{\sigma,\eta}\left\|\delta_\sigma*P-\delta_\eta*P\right\|_{\mathrm{TV}}=1-\min_{\sigma,\eta}\sum_{\tau}P(\sigma,\tau)\wedge P(\eta,\tau)$$

- Block criterion of weak ergodicity:

$$\sum_{n=1}^{\infty}\Big(1-\delta(P^{\mathrm{SCA}}_{\beta_n,q})\Big)=\infty \qquad \Rightarrow \qquad \delta(P^{\mathrm{SCA}}_{[\beta_\ell,\beta_n],q})\xrightarrow[n\uparrow\infty]{}0.$$

- The left is easy to show because

$$P^{\mathrm{SCA}}_{\beta_n,q}(\sigma,\tau)\ge\prod_x\frac{1}{1+e^{\beta_n|\tilde{h}_x(\sigma)+q_x\sigma_x|}}\ge\prod_x\frac{e^{-\beta_n\Gamma_x}}{2}=\frac{n^{-1}}{2^{|V|}}.$$

- For (ii), let $m=\min_{\sigma,\eta}\tilde{H}(\sigma,\eta)$ and

$$\mu_{\beta,q}(\sigma,\eta)=\frac{e^{-\beta\tilde{H}(\sigma,\eta)}}{\sum_{\sigma,\eta}e^{-\beta\tilde{H}(\sigma,\eta)}}\overset{\min q_x\ge\frac{\lambda}{2}}{=}\frac{e^{-\beta(\tilde{H}(\sigma,\eta)-m)}}{|\mathrm{GS}|+\sum_{\sigma,\eta:\tilde{H}(\sigma,\eta)>m}e^{-\beta(\tilde{H}(\sigma,\eta)-m)}}\xrightarrow[\beta\uparrow\infty]{}\pi^{\mathrm{G}}_\infty(\sigma)\,\delta_{\sigma,\eta}.$$

Since $$\frac{\partial\mu_{\beta,q}(\sigma,\eta)}{\partial\beta}=\Big(\underbrace{\mathbb{E}_{\mu_{\beta,q}}[\tilde{H}]}_{\xrightarrow[\beta\uparrow\infty]{}m}-\tilde{H}(\sigma,\eta)\Big)\mu_{\beta,q}(\sigma,\eta)\begin{cases}>0 & {}^\forall\beta<\infty \text{ if } \tilde{H}(\sigma,\eta)=m,\\ <0 & \text{if } \beta\gg1 \;\&\; \tilde{H}(\sigma,\eta)>m,\end{cases}$$

we can show $\sum_{n=N}^{\infty}\|\pi^{\mathrm{SCA}}_{\beta_{n+1}}-\pi^{\mathrm{SCA}}_{\beta_n}\|_{\mathrm{TV}}\le\frac{3}{2}$ for $N\gg1$

- Isolate the effect of $\boldsymbol{q} = \{q_x\}_{x\in V}$ as

$$P^{\text{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \prod_x \frac{e^{\frac{\beta}{2}q_x\sigma_x\eta_x}\cosh(\frac{\beta}{2}\tilde{h}_x(\boldsymbol{\sigma}))}{\cosh(\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x))}\frac{e^{\frac{\beta}{2}\tilde{h}_x(\boldsymbol{\sigma})\eta_x}}{2\cosh(\frac{\beta}{2}\tilde{h}_x(\boldsymbol{\sigma}))}$$
$$= \prod_{x\in D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} \varepsilon_x(\boldsymbol{\sigma})p_x(\boldsymbol{\sigma}) \prod_{y\in V\setminus D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} \Big(1-\varepsilon_y(\boldsymbol{\sigma})p_y(\boldsymbol{\sigma})\Big),$$

where $\varepsilon_x(\boldsymbol{\sigma}) = O_{\boldsymbol{\sigma}}(e^{-\beta q_x})$ and $D_{\boldsymbol{\sigma},\boldsymbol{\eta}} = \{x\in V : \sigma_x \neq \eta_x\}$.

- ε-SCA: $\varepsilon_x(\boldsymbol{\sigma})$ is uniformly replaced by a temperature-free $\varepsilon\in[0,1]$.

$$P^{\varepsilon\text{-SCA}}_{\beta}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \prod_{x\in D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} \varepsilon p_x(\boldsymbol{\sigma}) \prod_{y\in V\setminus D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} \Big(1-\varepsilon p_x(\boldsymbol{\sigma})\Big)$$
$$= \sum_{S:D_{\boldsymbol{\sigma},\boldsymbol{\eta}}\subset S\subset V} \varepsilon^{|S|}(1-\varepsilon)^{|V\setminus S|} \prod_{x\in D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} p_x(\boldsymbol{\sigma}) \prod_{y\in S\setminus D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} \Big(1-p_x(\boldsymbol{\sigma})\Big).$$

  - Initially motivated to reduce the memory size for generating random numbers.
  - Pros: the pinning effect is lighter, especially in the low-temperature regime.
  - Cons: there is no theoretical justification so far, an equilibrium measure is unknown, etc.

- Isolate the effect of $\boldsymbol{q} = \{q_x\}_{x\in V}$ as

$$P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \prod_x \frac{e^{\frac{\beta}{2}q_x\sigma_x\eta_x}\cosh(\frac{\beta}{2}\tilde{h}_x(\boldsymbol{\sigma}))}{\cosh(\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x))} \frac{e^{\frac{\beta}{2}\tilde{h}_x(\boldsymbol{\sigma})\eta_x}}{2\cosh(\frac{\beta}{2}\tilde{h}_x(\boldsymbol{\sigma}))}$$
$$= \prod_{x\in D_{\sigma,\eta}} \varepsilon_x(\boldsymbol{\sigma})p_x(\boldsymbol{\sigma}) \prod_{y\in V\setminus D_{\sigma,\eta}} \Big(1-\varepsilon_y(\boldsymbol{\sigma})p_y(\boldsymbol{\sigma})\Big),$$

where $\varepsilon_x(\boldsymbol{\sigma}) = O_{\boldsymbol{\sigma}}(e^{-\beta q_x})$ and $D_{\boldsymbol{\sigma},\boldsymbol{\eta}} = \{x \in V : \sigma_x \neq \eta_x\}$.

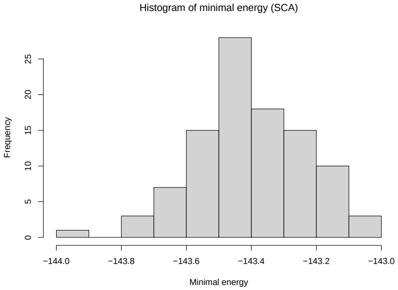
- ε-SCA: $\varepsilon_x(\boldsymbol{\sigma})$ is uniformly replaced by a temperature-free $\varepsilon \in [0, 1]$.

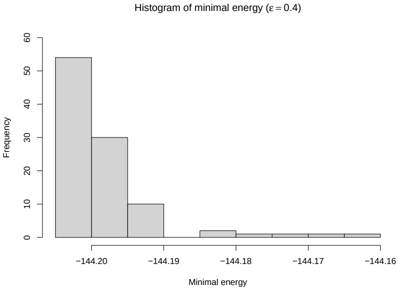
$$P^{\varepsilon\text{-SCA}}_{\beta}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \prod_{x\in D_{\sigma,\eta}} \varepsilon p_x(\boldsymbol{\sigma}) \prod_{y\in V\setminus D_{\sigma,\eta}} \Big(1-\varepsilon p_x(\boldsymbol{\sigma})\Big)$$
$$= \sum_{S:D_{\sigma,\eta}\subset S\subset V} \varepsilon^{|S|}(1-\varepsilon)^{|V\setminus S|} \prod_{x\in D_{\sigma,\eta}} p_x(\boldsymbol{\sigma}) \prod_{y\in S\setminus D_{\sigma,\eta}} \Big(1-p_x(\boldsymbol{\sigma})\Big).$$

  - Initially motivated to reduce the memory size for generating random numbers.
  - Pros: the pinning effect is lighter, especially in the low-temperature regime.
  - Cons: there is no theoretical justification so far, an equilibrium measure is unknown, etc.

Application to the TSP (Fukushima-Kimura):

- $|V| = 10$, $\{d_{x,y}\}_{V\times V}$: i.i.d. uniform on $[0,1]$.
- $\beta_n = \beta_0 e^{\alpha n}$ (exponential cooling) with $\alpha = 0.0001$, $\beta_0 = 0.001$.
- 100 samples, a 120000-step MC each.





- SCA: the minimum-energy ($= -143.9403$) spin configuration is not legitimate.
- ε-SCA with $\varepsilon = 0.4$: the minimum-energy ($= -144.2006$, success rate 54%) spin configuration is legitimate.

- What we wanted: faster dynamics based on multi-spin updates.
- What we proposed: use the stochastic cellular automata (SCA):

$$P^{\text{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \prod_x \frac{e^{\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x)\eta_x}}{2\cosh(\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x))} = \prod_{x\in D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} \varepsilon_x(\boldsymbol{\sigma})p_x(\boldsymbol{\sigma}) \prod_{y\in V\setminus D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} \Big(1-\varepsilon_y(\boldsymbol{\sigma})p_y(\boldsymbol{\sigma})\Big).$$

- What we have shown:
  - Mixing is much faster than Glauber in the high-temperature regime:

$$r = \max_x \left(\tanh\frac{\beta q_x}{2} + \sum_y \tanh\frac{\beta|J_{x,y}|}{2}\right) < 1 \quad\Rightarrow\quad T^{\text{SCA}}_{\text{mix}}(\tfrac{1}{e}) \le \frac{\log|V|+1}{\log(1/r)}.$$

  - A standard cooling schedule works for the SCA, too, to find a $\boldsymbol{\sigma}_{\text{GS}}$:

$$\beta_n = \frac{\log n}{\sum_x(q_x+|h_x|+\sum_y|J_{x,y}|)} \quad\Rightarrow\quad \sup_\mu \Big\|\mu * P^{\text{SCA}}_{\beta_1,\boldsymbol{q}} * \cdots * P^{\text{SCA}}_{\beta_n,\boldsymbol{q}} - \text{Unif(GSs)}\Big\|_{\text{TV}} \xrightarrow[n\uparrow\infty]{} 0.$$

- Ongoing work (supported by numerical evidence):
  - ε-SCA, where the pinning parameters $\boldsymbol{q}=\{q_x\}_{x\in V}$ are turned off and a collection of spins is chosen by Binom$(|V|,\varepsilon)$:

$$P^{\varepsilon\text{-SCA}}_\beta(\boldsymbol{\sigma},\boldsymbol{\eta}) = \sum_{S:D_{\boldsymbol{\sigma},\boldsymbol{\eta}}\subset S\subset V} \varepsilon^{|S|}(1-\varepsilon)^{|V\setminus S|} \prod_{x\in D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} p_x(\boldsymbol{\sigma}) \prod_{y\in S\setminus D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} \Big(1-p_x(\boldsymbol{\sigma})\Big).$$

  - First hit to the target states, under an exponential cooling schedule.
    (First hit to the target states ↔ convergence in total variation; exponential ↔ logarithmic)

- What we wanted: faster dynamics based on multi-spin updates.
- What we proposed: use the stochastic cellular automata (SCA):

$$P^{\mathrm{SCA}}_{\beta,\boldsymbol{q}}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \prod_x \frac{e^{\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x)\eta_x}}{2\cosh(\frac{\beta}{2}(\tilde{h}_x(\boldsymbol{\sigma})+q_x\sigma_x))} = \prod_{x\in D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} \varepsilon_x(\boldsymbol{\sigma})p_x(\boldsymbol{\sigma}) \prod_{y\in V\setminus D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} \big(1-\varepsilon_y(\boldsymbol{\sigma})p_y(\boldsymbol{\sigma})\big).$$

- What we have shown:
  - Mixing is much faster than Glauber in the high-temperature regime:

$$r = \max_x \bigg(\tanh\frac{\beta q_x}{2} + \sum_y \tanh\frac{\beta|J_{x,y}|}{2}\bigg) < 1 \quad\Rightarrow\quad T^{\mathrm{SCA}}_{\mathrm{mix}}(\tfrac{1}{e}) \le \frac{\log|V|+1}{\log(1/r)}.$$

  - A standard cooling schedule works for the SCA, too, to find a $\boldsymbol{\sigma}_{\mathrm{GS}}$:

$$\beta_n = \frac{\log n}{\sum_x(q_x+|h_x|+\sum_y|J_{x,y}|)} \quad\Rightarrow\quad \sup_\mu \Big\|\mu * P^{\mathrm{SCA}}_{\beta_1,\boldsymbol{q}} * \cdots * P^{\mathrm{SCA}}_{\beta_n,\boldsymbol{q}} - \mathrm{Unif(GSs)}\Big\|_{\mathrm{TV}} \xrightarrow[n\uparrow\infty]{} 0.$$

- Ongoing work (supported by numerical evidence):
  - ε-SCA, where the pinning parameters $\boldsymbol{q} = \{q_x\}_{x\in V}$ are turned off and a collection of spins is chosen by $\mathrm{Binom}(|V|,\varepsilon)$:

$$P^{\varepsilon\text{-SCA}}_{\beta}(\boldsymbol{\sigma},\boldsymbol{\eta}) = \sum_{S:D_{\boldsymbol{\sigma},\boldsymbol{\eta}}\subset S\subset V} \varepsilon^{|S|}(1-\varepsilon)^{|V\setminus S|} \prod_{x\in D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} p_x(\boldsymbol{\sigma}) \prod_{y\in S\setminus D_{\boldsymbol{\sigma},\boldsymbol{\eta}}} \big(1-p_x(\boldsymbol{\sigma})\big).$$

  - First hit to the target states (↔ convergence in total variation), under an exponential (↔ logarithmic) cooling schedule.